Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 39(2), ss. 475-488, Haziran 2024
Cukurova University Journal of the Faculty of Engineering, 39(2), pp. 475-488, June 2024

Comparison of Random Forest and Support Vector Regression Models in Predicting Hydrogen Production Process from Biomass

Gülbahar BİLGİÇ TÜZEMEN*[1] ORCID *0000-0002-9503-5884*
Ali Emre GÖK[2] ORCID *0009-0009-8292-2132*

[1]Nevşehir Haci Bektas Veli University, Faculty of Engineering-Architecture, Department of Metallurgy and Material Engineering, Nevşehir, Türkiye
[2]Nevşehir Haci Bektas Veli University, Faculty of Engineering-Architecture, Department of Computer Engineering, Nevşehir, Türkiye

Geliş tarihi: 04.01.2024 Kabul tarihi: 27.06.2024

Atıf şekli/ How to cite: BİLGİÇ, G., GÖK, A.E., (2024). Comparison of Random Forest and Support Vector Regression Models in Predicting Hydrogen Production Process from Biomass. Cukurova University, Journal of the Faculty of Engineering, 39(2), 475-488.

Abstract

The need for energy in the world is increasing day by day and various energy production methods are used to meet this need. Production of hydrogen from biomass is one of these methods. Hydrogen production from biomass is a promising process to produce hydrogen and energy which has advantages such as the ability to use sustainable energy sources like biomass and solid waste, being carbon neutral, and increasing energy independence thanks to the variation of resources and the availability of local resources. The catalysts used in this process which can be conducted in three separate ways, affect hydrogen and energy production positively or negatively. One of the most important steps in effectively acquiring the ideal amount of product is predicting the outcomes of this procedure. This article compares a support vector regression (SVR) and random forest (RF) model to predict how various inputs used to produce hydrogen from biomass will affect hydrogen output. Additionally, the effect of catalyst addition on hydrogen yield in biomass processes was examined. In this context, 57 experimental studies from the literature were selected as a data set. From this data, 90% was selected for training and 10% for testing. The outputs were evaluated according to parameters such as R^2, RMSE and MSE. The results show that RF and SVR models can significantly predict catalyst activity and hydrogen production.

Keywords: Hydrogen, Biomass, Random forest, Support vector regression

Biyokütleden Hidrojen Üretiminde Rastgele Orman ve Destek Vektör Regresyon Modellerinin Kıyaslaması

Öz

Dünyadaki enerji ihtiyacı günden güne artış göstermekte ve bu ihtiyacın karşılanması için, çeşitli enerji üretim yöntemleri kullanılmaktadır. Bu yöntemlerden biri biyokütleden hidrojen üretimidir. Biyokütle ve atık benzeri yenilenebilir enerji kaynaklarından yararlanma kabiliyeti, karbon nötr olması, kaynak çeşitliliği

*Sorumlu yazar (Corresponding Author): Gülbahar BİLGİÇ TÜZEMEN, *glbhrblg@gmail.com*

ve yerel kaynakların kullanılabilirliği sayesinde enerji bağımsızlığını artırması gibi avantajları bulunan biyokütleden hidrojen üretimi, gelecek vaat eden bir hidrojen ve enerji üretim sürecidir. Üç farklı yöntem kullanılarak gerçekleştirilebilen bu süreçte kullanılan katalizörler, hidrojen üretimine olumlu ve olumsuz etki etmektedir. Bu sürecin sonuçlarını tahmin etmek, optimum miktarda ürünü verimli bir şekilde elde etmede kritik bir adımdır. Bu makalede, biyokütleden hidrojen üretmek için kullanılan çeşitli girdilerin hidrojen çıktısını nasıl etkileyeceğini tahmin etmek için bir destek vektör regresyonunu (SVR) ve rastgele orman (RF) modelini karşılaştırılmıştır. Ayrıca biyokütle süreçlerinde katalizör ilavesinin hidrojen verimi üzerinde etkisini incelenmiştir. Bu bağlamda literatürden 57 deneysel çalışma veri seti olarak seçilmiştir. Bu verilerden eğitim için %90 ve test için %10 seçilmiştir. Sonuçlar R2, RMSE ve MSE gibi parametrelere göre değerlendirilmiştir. Sonuç olarak RF ve SVR modellerinin katalizör aktivitesini ve hidrojen üretimini önemli ölçüde tahmin edebildiğini göstermektedir.

Anahtar Kelimeler: Hidrojen, Biyokütle, Rastgele orman, Destek vektör regresyonu

1. INTRODUCTION

Nowadays, the majority of H_2 is produced commercially using inefficient and unsustainable methods like coal gasification and methane vapor reforming. Using tried-and-true thermochemical processes like pyrolysis, liquefaction, and gasification, carbonaceous materials may be converted into H_2-rich syngas and other hydrocarbons [1].

The effective thermochemical conversion process called gasification may combine solid biomass into flammable gases, such as carbon monoxide, carbon dioxide, hydrogen, methane, light hydrocarbons, and coal. Solid biomass may then be converted into electrical energy using this procedure. The catalytic gasification of municipal solid waste (MSW) is one of the best methods for producing syngas, or hydrogen-rich gas, from renewable sources [2]. This gas is utilized as feedstock for fuel cell applications and hydrogen combustion engines to release stored energy, or it can be used to create hydrogen to synthesize methanol and ammonia [3]. Moreover, because of the easy conditions and lower temperatures, it also lessens the creation of thermal NOx [2].

Among thermochemical processes, gasification is seen as the most promising because of its high conversion efficiency and ability to manage various kinds of waste. Nonetheless, the tar produced during gasification is one of the primary issues. Tar content tends to clog pipelines, auxiliary equipment, and heat exchanger tubes, which raises maintenance costs and reduces process efficiency overall. Significant efforts have been made to remove tar from fuel gas. Thermal processes such as catalytic cracking and high-temperature cracking have been the primary methods used so far. Because catalytic cracking can operate at comparatively lower temperatures and has high tar removal effectiveness, it is believed to be the most effective technology for minimizing tar production in the gas mixture. Due to their unique features in a variety of sectors, nanomaterials have garnered a lot of interest recently as catalysts in comparison to their bulk counterparts.

Many researchers have reported on various catalysts for gasification and pyrolysis processes, each with a different promoter and support [4,5]. Dolomite, nickel, dolomite, and olivine have been extensively used as catalysts in biomass gasification due to their affordability, availability, and ability to significantly reduce tar content in the gasifier effluent [2,6,7].

When paired with transition metal promoters, a catalyst may be used in a variety of ways to catalyze the conversion of MSW. For example, the impact of incorporating transition metal-containing Ni-CaO-based catalysts into Ni-CaO-based catalysts (such as Fe, Cu, Co, and Zn) for gasification and pyrolysis processes has been studied [5,8]. This can improve the process's energy density and efficiency while also raising the total production of syngas and hydrogen [9]. Also, the catalytic characteristics of nanoscale NiO (nano-NiO) particles have garnered significant interest. More precisely, nano-NiO

particles may be applied to the surface of different carriers (such as alumina and Al_2O_3) to produce the supported catalyst at a lower cost [10].

Recently, as in many different fields, there has been a great interest in machine learning (ML) to optimize the operating conditions of the gasification system. Regression analysis, artificial neural networks (ANN) [11,12], tree-based approaches like classification and random forest (RF) [13] based on regression trees (RT) and extreme trees, and support vector machines (SVM) [14] are frequently supported when modeling hydrogen generation processes from biomass [15].

Random forest, first introduced by Breiman [16], is a machine-learning technique that has become popular for prediction, analyzing variable importance, outlier identification, classification, and variable selection. The RF model uses a group of decision trees and an average of their estimations to solve classification or regression problems effectively. The branches of a tree are visited according to a set of constraints and relationships from root to leaf, and each tree selects the most popular one from its branches. These forecasts of related trees are averaged to acquire estimation. Input handling without entries, a smaller number of parameters high accuracy and noise resistance are some of the advantages of RF [16].

Decision trees in RF improve the efficiency of optimization models thanks to their high diversity and low bias. Furthermore, RF has a lower overfitting risk, and it has a hyperparameter for determining the number of estimators (trees) in it. In Figure 1, the schema of the RF algorithm is given. As seen in Figure 1, input data is presented to all the decision trees in the RF. From top to bottom, the nodes of a decision tree are called root, node, and leaf. The root node does not have parent nodes, while leaves do not have any children [17].

A support vector regressor (SVR) is an SVM-based model that is specialized for regression problems. SVR is distinguished by its kernels, number of support vectors, margin control, and sparse solutions. The regression of the SVR is like the regression of neural networks. The main difference is that in the SVR, input layer weights are a subset of training samples. Even though SVR is less known than SVM, it has been a more effective tool in value estimation. In the training phase, it equally deals with high and low estimations that are wrongly done. In the SVR, the complexity of computations does not change according to input space; this is one of the benefits of the SVR. Besides, the SVR predicts highly accurately and is successful in generalization [18,19]. The schema of the SVR is given in Figure 2.

Mathematical models of the catalytic gasifier system can help cut down on the cost and duration of labor-intensive trials. However, gasification and pyrolysis are complex processes that mostly occur because of overlapping interactions, including solid-gas and gas-gas [19]. In the literature, many researchers have reported on various RF, SVR, and hybrid models for gasification and pyrolysis (the production of hydrogen-rich gas with and without catalysts) processes.

Mutlu, A.Y. and Yucel [20] employed RF and SVM-based classifiers to forecast the composition of syngas emitted during the gasification process because of their capacity to discern subtle patterns in noisy and complicated datasets. Both binary and multi-class classifiers, such as least squares SVM and RF, are trained using calorific values with discrete levels, yielding classification accuracy of over 89% and 96%, respectively. Two types of classifiers have been developed to anticipate the calorific value and composition of the producer gas generated in a downdraft gasifier during the gasification of woody biomass. The proposed methods were created and evaluated using 10-fold cross-validation on 5237 data samples. Binary and multi-class classifiers yielded prediction accuracy rates of more than 96% and 89%, respectively.

Elmaz, F. et. al. [21] assessed regression techniques (SVR, DTR, PR, and MLP) to predict the CH_4, H, CO, and heating value (HHV) outputs of the downdraft biomass gasification process using an experimentally acquired data set. For HHV determination coefficient (R^2)>85 outputs and outputs, respectively, MLP and DTR were obtained, and they beat the other methods. Quadratic PR had

the worst performance of all the recommended techniques. With R^2>0.9 for the majority of outcomes, multilayer perceptron and decision tree regression fared better than the other modeling approaches.

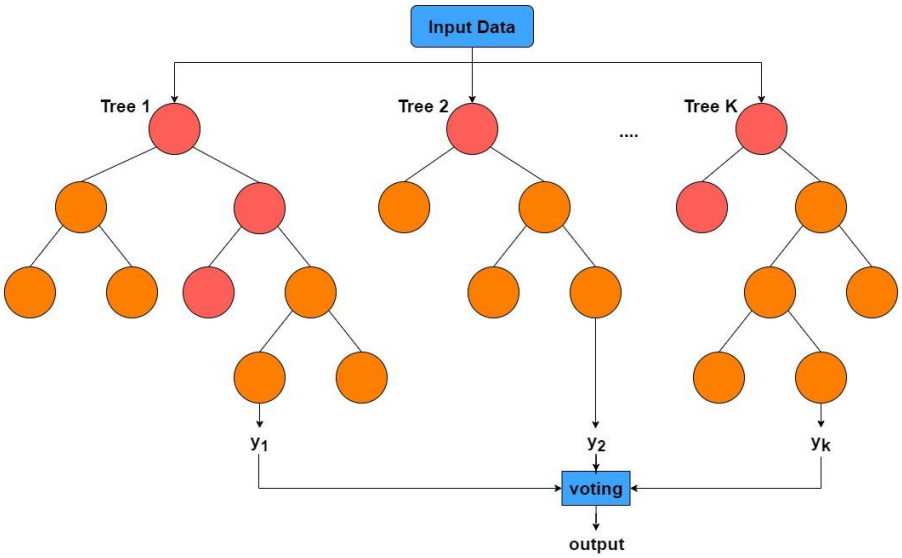

Figure 1. The random forest algorithm

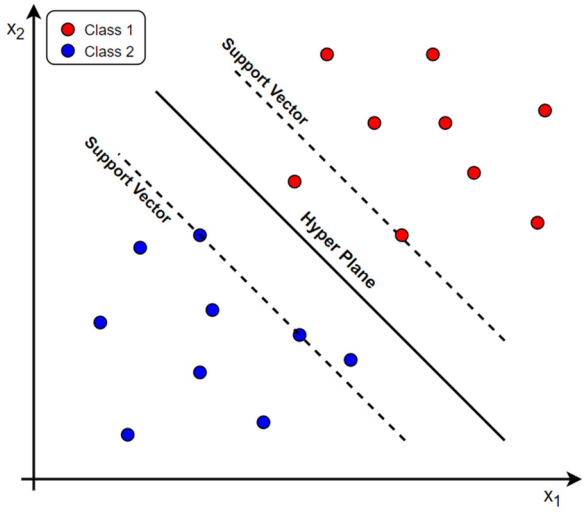

Figure 2. The support vector regression

Leng et. al. [41] developed RF models with high accuracy and outstanding generalization capacity to compare both pyrolysis conditions and comprehensive raw material parameters. In this work, gradient boosting, RF-based regression prediction models for three-phase product distribution and bio-oil HHV were built by collecting experimental data of rapid pyrolysis of lignocellulosic biomass in bubbly fluidized bed from earlier literature. As input attributes,

comprehensive raw material properties and pyrolysis circumstances were considered and compared. With its high accuracy and outstanding generalization ability, the RF has been demonstrated to be the most appropriate algorithm for predicting three-phase product yields and bio-oil HHV among other algorithms.

Xing J et. al. [23] used an RF model to forecast the primary chemical components of biomass based on the final analysis. Two datasets were constructed for the RF model's training and application using previously published information. The application findings demonstrate that the current RF model can predict chemical components for a variety of biomasses with high accuracy; the predicted values for lignin, hemicellulose, and cellulose are 0.962, 0.904, and 0.862, respectively.

Considering this information, the RF and SVR models were chosen to optimize the conditions for hydrogen production processes from biomass and to predict the effect of the catalyst. To collect information for data set preparation, literature studies on hydrogen production by gasification or pyrolysis were evaluated. In this context, 57 published data samples have been collected from experimental studies conducted in the literature. Data from 57 datasets were applied to build and evaluate ML models. As an information preparation process, the inputs are approximate analysis of biomass waste feedstocks, their volatile matter (VM), fixed carbon (FC), hydrogen (H), sulfur (S), carbon (C), nitrogen (N), and oxygen (O), temperature and the name of the catalyst were determined as. Only hydrogen production was considered as output. These examples were divided into training and application types, at rates of 90% and 10%, respectively, as in previous machine learning studies. To our knowledge, this is the first study in the literature on the effect of catalyst use.

In the second stage of this study, the dataset, preprocessing methods, and model evaluation metrics were explained. Afterwards, at the third stage, results obtained from this study were given with related graphics. Finally, in the fourth chapter, a summary of this study and plans for further research were given.

2. MATERIALS AND METHODOLOGY

2.1. Sample Experimental Dataset Collection

Literature studies on hydrogen obtained from pyrolysis and gasification experiments were evaluated for data set preparation. In the created dataset, MSW samples include kitchen garbage, paper, textiles, plastic, dried sewage sludge, cassava shells, pine sawdust, corn stalks, sawdust, etc. Different biomasses were used. According to earlier machine learning research, such samples are split into training and test types with proportions of 90% and 10%, respectively. In this case, the RF model's inputs are the carbon percent, hydrogen-carbon ratio, and oxygen-carbon ratio; the output is the hydrogen yield. Figure 3 is the visualization of the use of machine learning algorithms for hydrogen output prediction by input and output. Table 1 describes the independent variables of the data set, and Table 2 shows examples of datasets. The data were collected from the sources specified in the literature [2,8,24-35]. Data in which catalysts were not used in production processes where hydrogen was released was also included in the study.

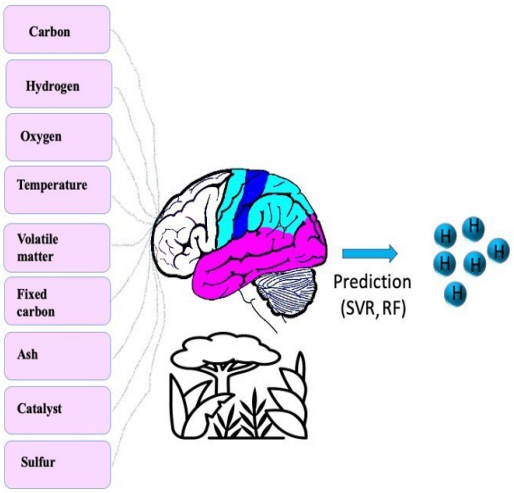

Figure 3. ML algorithms for hydrogen prediction approaches

Table 1. The description of independent variables of input and output

		Symbol	Unit
Input	Carbon	C	wt%
	Hydrogen	H	wt%
	Oxygen	O	wt%
	Nitrogen	N	wt%
	Sulphur	S	wt%
	Volatile matter	VM	wt%
	Fixed carbon	FC	wt%
	Ash		wt%
	Catalyst	Ni-Fe-WMP CaO NiO/MD Ni-Zn-WMP etc.	-
	Temperature	°C	Celsius
Output	Hydrogen Yield	HY	(vol.%)

Table 2. Sample of data set

C	H	O	N	S	VM	FC	Ash	Catalyst	Tem.	HY
52.2	5.5	40	2.3	0	82.7	12	5.3	Nickel based	850	46.5
51.81	5.76	30.22	0.26	0.36	82.28	11.79	5.93	$CaMg(CO_3)_2$	700	27.01
42.11	5.33	37.08	1.42	0.11	73.62	12.43	10.5	CaO	650	61.23
50.48	5.92	42.14	0.27	0.02	75.6	22.9	1.5	without catalyst	848	15.5
50.26	6.72	42.66	0.16	0.2	77.71	16.94	0.34	CaO	650	46.6
42.11	5.33	37.08	1.42	0.11	73.62	12.43	10.5	CaO	650	58.69
46.36	5.75	43.62	2.26	0.32	82.18	16.13	1.69	Nickel based	900	52.3
52.2	5.5	40	2.3	0	82.7	12	26.14	Nickel based	850	46.3
50.26	6.72	42.66	0.16	0.2	77.71	16.94	0.34	CaO	700	56.1
50.26	6.72	42.66	0.16	0.2	77.71	16.94	0.34	CaO	800	53.8
52.2	5.5	40	2.3	0	82.7	12	26.14	Nickel based	850	47.8
42.11	5.33	37.08	1.42	0.11	73.62	12.43	10.5	CaO	650	60.28
51.81	5.76	30.22	0.26	0.36	82.28	11.79	5.93	CaO	900	48.63
42.54	5.86	31.35	1.07	0.2	64.19	16.87	9.26	CaO	700	29.8
49.74	7.82	41.6	0.82	0.05	86.4	10.61	4.77	without catalyst	1000	51.36
40.31	5.29	21.21	5.73	1.32	61.56	7.01	26.14	without catalyst	793	45.35
51.81	5.76	30.22	0.26	0.36	82.28	11.79	5.93	CaO	750	34.7
47.1	5.9	46.9	0.1	0	74.6	18.1	1.1	Nickel based	800	49
50.26	6.72	42.66	0.16	0.2	77.71	16.94	0.34	CaO	750	60.8

Both RF and SVR models have hyperparameters that affect model performance positively and negatively. The RF has hyperparameters such as the number of estimators (number of trees), criterion, max depth, min samples split, min samples leaf, max features, bootstrap, and random state. Cost (C), kernel, epsilon, gamma, degree and shrinking are hyperparameters used in the SVR model. In this study, the number of estimators, min samples split, and min samples leaf hyperparameters in the RF model and in the SVR model, C, kernel, and epsilon hyperparameters were used as presented in Table 3.

Table 3. Tuning parameters and values of applied methods

Method	Tuning parameter	Value
RF	Number of estimators	10
	Min samples split	2
	Min samples leaf	1
SVR	Cost (C)	2.811
	Kernel	linear
	Epsilon	0.67

2.2. Feature Extraction and Preprocessing

Once data collection was completed, some data pre-processing techniques were applied to the dataset to improve the models' performance. Since ML and deep learning models struggle to work with categorical data, they must be presented numerically. Therefore, the one-hot encoding method was applied to the dataset. This method transforms every categorical data into a column, and the presence and absence are respectively indicated with 1 and 0 values. Thanks to one-hot encoding, models understand the dataset better and perform better [12]. Another pre-processing technique that was applied to the dataset was normalization. Normalization is the scaling of data to a certain general range or distribution. That allows comparison between different data sets and improves performance. Besides, it reduces the complexity of the calculation process and calculation time. In this study, the dataset was normalized to a (-1, 1) interval by using the corresponding Python library [36]. The dataset was split into training and test sets once pre-processing was finished. The remaining 90% of the data was utilized in the test set for model evaluation, with the training set being used for model training.

2.3 Evaluation

Following model development, the model accuracy and performance were assessed. The models' performance was assessed using the determination coefficient (R^2), variance, mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) [2].

In our example, the RF and SVR models include storage data obtained from the biomass gasification and pyrolysis methods. As part of the information preparation process, the inputs were determined as the name of biomass waste feedstocks, VM, FC, C, H, N, S, O, temperature, and catalyst. The output/targeted variable was selected as H_2. For the training data, a total of 52 runs (or 90% of the data) and 5 runs (10% of the data) were chosen by [37]. R^2, RSME, MAE and MSE, represented by the equations in Table 4 respectively, were used for the model's accuracy values.

Table 4. Evaluation indicators of models' prediction performance

Method	Mathematical expression
Mean absolute error	$\frac{1}{n}\Sigma_{i=1}^{n}\lvert f_i - y_i\rvert$
Mean square error	$\frac{1}{n}\Sigma_{i=1}^{n}(f_i - y_i)^2$
Root mean square error	$\sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(f_i - y_i)^2}$
Determination coefficient	$1 - \Sigma_{i=1}^{n}(y_i - f_i)^2 / \Sigma_{i=1}^{n}\left(y_i - \underline{y}\right)^2$

3. RESULT AND DISCUSSION

3.1. Performance Evaluation of Test Results

Table 5 shows the results of the two models' ML analyses (RF and SVR). The models used in this study were evaluated separately for both train and test sets based on MAE, MSE, RMSE, and R^2. The values of each metric are given in Table 5 and Figure 4. The data that are divided into test and training data at different rates have a coefficient of determination that ranges from 0.57 to 0.98, according to the results. R^2 was found to be 0.98 in the RF model and 0.94 in the SVR model. For the test data in the RF model, the minimal values of the criterion MAE, MSE, and RMSE were determined

to be 1.00, 1.31, and 1.15, respectively. For the test data in the SVR model as well, the minimum values of the criterion MAE, MSE, and RMSE were determined to be 1.34, 3.67, and 1.92, respectively. For the training data in the RF model, it was discovered that the minimal values of the criterion MAE, MSE, and RMSE were 2.64, 12.00, and 3.47, respectively These results indicate that the RF model is more performant than the SVR model.

Figure 4 shows the hydrogen yield prediction performances of the RF (a) and SVR (b) models on the training set. The expected values are shown with the red line and predictions are shown with stars The RF model's training RMSE and R^2 values were obtained as 3.47 and 0.93, respectively, while these values were obtained as 1.15 and 0.98 in the test dataset. These values proved its dependability in forecasting unknown data. The performance of the SVR model in the test set (R^2 = 0.94 and RMSE = 1.92) showed that the model was good at forecasting unknown data. However, the SVR model's deficient performance on the training dataset, with an R^2 = 0.57 and an RMSE = 8.80, showed its low predictive ability with known data. In Figures 5 and 6, prediction performances of the RF and SVR models in the test dataset are given. While real values are shown with the red line, estimations of the models are shown with circles.

The relative error distribution of the models is given in Figure 7. Analysis of relative error distribution is used for analyzing the reliability and performance of the models. It helps to understand how accurate or inaccurate the predictions are and to improve the models accordingly. In Figure 7, The blue and orange dots show the test and train error distribution of the RF model, while the green and red dots show the test and train error distribution of the SVR model, respectively.

Table 5. Analysis of train and test set outcomes of models

Method	Dataset	MAE	MSE	RMSE
RF	Train	2.64	12.00	3.47
	Test	1.00	1.31	1.15
SVR	Train	6.35	77.42	8.80
	Test	1.34	3.67	1.92

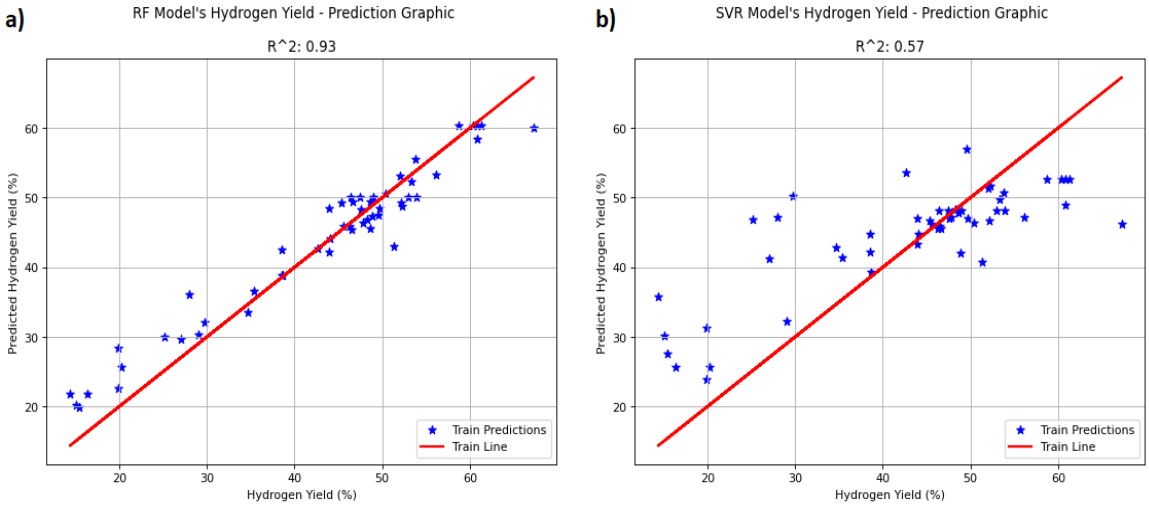

Figure 4. Hydrogen yield prediction performances of random forest (a) and support vector regressor (b) models on the training dataset

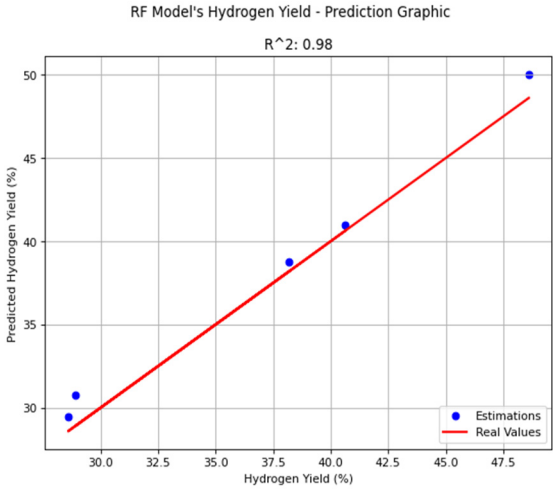

Figure 5. Prediction performance and R^2 score of random forest model

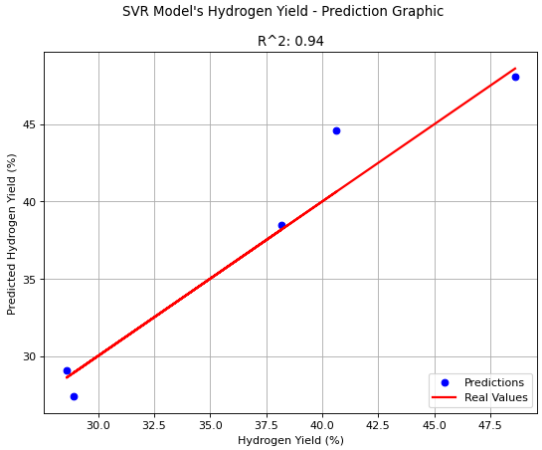

Figure 6. Prediction performance and R^2 score of support vector regression model

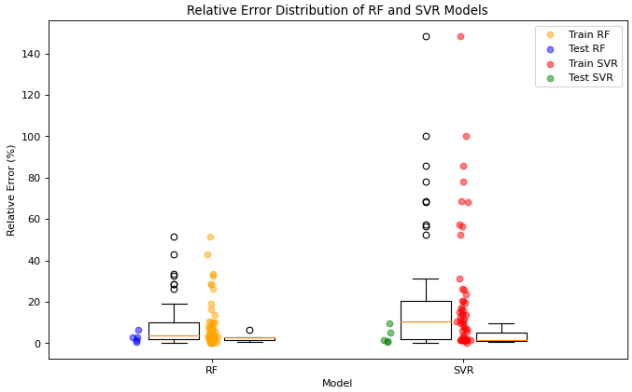

Figure 7. Relative error distribution of the models

3.2. Sensitivity Analysis

When assessing the degree of relationship between two variables, the Pearson correlation coefficient is commonly employed. A Pearson correlation coefficient (r) matrix is displayed in Figure 8, with each row denoting a variable from the dataset and each column representing the same variables [38]. The correlation coefficient between two variables is shown in each cell. The hydrogen generation is shown to be significantly influenced by the feedstock and experimental operating conditions, as indicated by the Pearson correlation coefficient matrix. The high positive correlations between feedstock attributes show that the choice of feedstock has a major effect on hydrogen generation. The positive relationship between carbon, hydrogen, volatile mixtures, and fixed carbon mixtures shows that the composition of the feedstock has a direct effect on the hydrogen yield. In contrast, the mixture of nitrogen, sulfur, and temperature has a negative relationship with pyrolytic product yields [15]. Since each feature contributes separately to the model, the absence of significant correlations between multiple input features helps preserve all these features. The negative correlation in both trait values indicated that these products reduced yield. Other attributes, on the other hand, showed fewer effects on hydrogen yield, with absolute correlation values less than 0.15. For instance, there was a 0.49 correlation coefficient between the carbon and oxygen levels of biomass characteristics. This result was in line with earlier experimental research [39].

The significance of specifying input variables in the hydrogen output of the RF model for the biomass process is illustrated in Figure 9. It is evident that the catalyst is a key factor in the predicted generated gas (26 % together), while it is also important for temperature (14%), H (16 %), N (7 %), FC (6 %), VM (6 %), ash (6 %) and less important for C (7 %), O (0.0256 %) and S (5%). Therefore, this study has proven that the catalyst has a powerful, positive effect on HP.

4. CONCLUSIONS AND FURTHER RESEARCH

Because biomass has a high conversion efficiency, gasification has drawn interest from academic and industry researchers. Nonetheless, the primary obstacle to the advancement of biomass gasification is tar production. Researchers have employed different catalysts to break down tar and create H_2-rich gas. Machine learning can help the process, though, as there is still a dearth of material in this field in the literature. Based on earlier research, the best catalyst to employ for this purpose was identified. This work uses SVR and RF algorithms to anticipate hydrogen generation. The machine learning evaluation accurately predicted the sort of catalyst that would be utilized to produce hydrogen from waste biomass. In addition, an assessment of each variable's relative relevance was made. For the RF and SVR methods, the R^2 values were 0.98 and 0.94, respectively. From the modeling results, it was seen that nickel-based and calcium-oxide catalysts would be more suitable. Furthermore, the training phase yields accuracies of 1.15 to 1.92 of RMSE, and 1.00 to 1.34 of MAE. It was evident from the modeling findings that catalysts based on calcium oxide and nickel would be more appropriate.

According to our knowledge, this is the first study in the literature on the effect of catalyst use. Additionally acquiring experimental data is a costly and prolonged process. Therefore, the dataset used in this study is limited. This limitation has a negative impact on the training and generalization abilities of the models. In further research, better results may be obtained by collecting a dataset that has more samples and a balanced distribution. In addition, the models have some limitations. The SVR model has limitations such as the determination of the kernel function, hyperparameter adjustment and data scaling. The SVR model's performance is dependent on its kernel function and hyperparameters. Determining the right type of kernel function and finding the best hyperparameters are crucial while using that model. The SVR model is also affected by the data scaling.

Improper scaling data may affect the performance negatively. The RF model has limitations such as the high cost of computation and the long processing time.

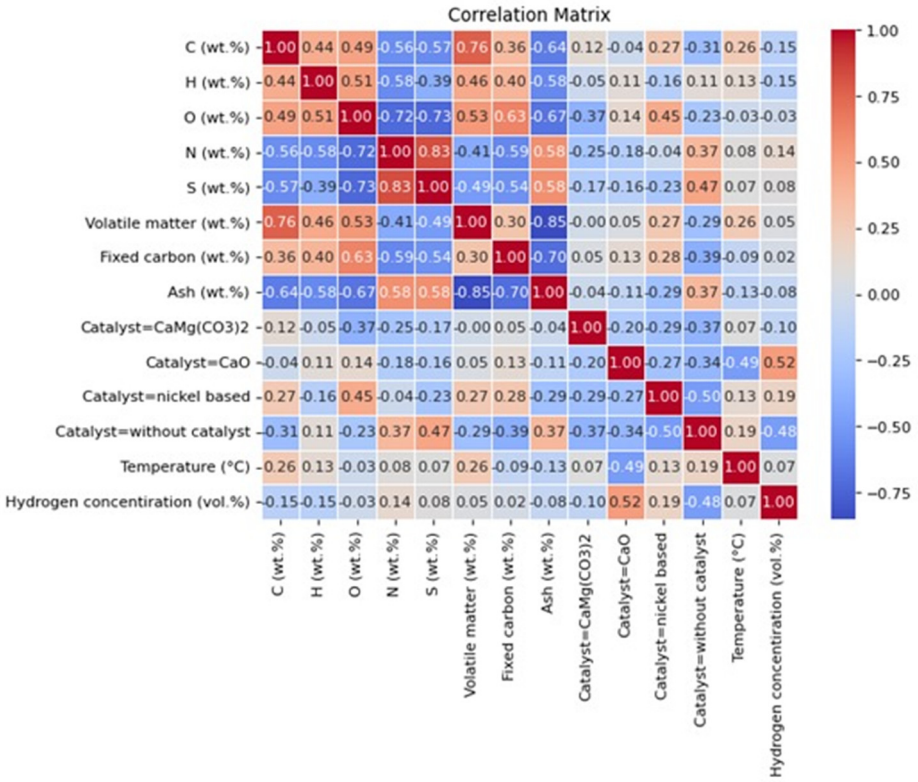

Figure 8. Correlation between parameters

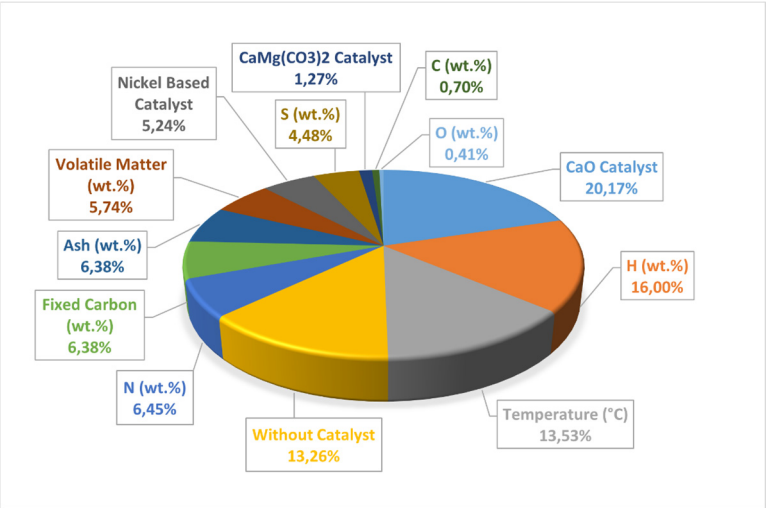

Figure 9. Feature importance (%) of input variables on hydrogen output of RF model for Biomass process

5. REFERENCES

1. Jamro, I.A., Raheem, A., Khoso, S., Baloch, H.A., Kumar, A., Chen, G., Bhagat, W.A., Wenga, T., Ma, W., 2023. Investigation of Enhanced H2 Production from Municipal Solid Waste Gasification Via Artificial Neural Network with Data on Tar Compounds. Journal of Environmental Management, 328, 117014.

2. He, M., Hu, Z., Xiao, B., Li, J., Guo, X., Luo, S., Yang, F., Feng, Y., Yang, G., Liu, S., 2009. Hydrogen-rich Gas from Catalytic Steam Gasification of Municipal Solid Waste (MSW): Influence of Catalyst and Temperature on Yield and Product Composition. International Journal of Hydrogen Energy, 34(1), 195-203.

3. Wu, M.-H., Lin, C.-L., Zeng, W.-Y., 2014. Effect of Waste Incineration and Gasification Processes on Heavy Metal Distribution. Fuel Processing Technology, 125, 67-72.

4. Gao, N., Liu, S., Han, Y., Xing, C., Li, A., 2015. Steam Reforming of Biomass Tar for Hydrogen Production over NIO/Ceramic Foam Catalyst. International Journal of Hydrogen Energy, 40(25), 7983-7990.

5. Irfan, M., Li, A., Zhang, L., Javid, M., Wang, M., Khushk, S., 2019. Enhanced H2 Production from Municipal Solid Waste Gasification Using Ni–Cao–Tio2 Bifunctional Catalyst Prepared by DC Arc Plasma Melting. Industrial & Engineering Chemistry Research, 58(29), 13408-13419.

6. Zhou, C., Yrjas, P., Engvall, K., 2021. Reaction Mechanisms for H2O-Enhanced Dolomite Calcination at High Pressure. Fuel Processing Technology, 217, 106830.

7. Soomro, A., Chen, S., Ma, S., Xiang, W., 2018. Catalytic Activities of Nickel, Dolomite, and Olivine for Tar Removal and H2-Enriched Gas Production in Biomass Gasification Process. Energy & Environment, 29(6), 839-867.

8. Irfan, M., Li, A., Zhang, L., Ji, G., Gao, Y., Khushk, S., 2021. Hydrogen-rich Syngas from Wet Municipal Solid Waste Gasification Using Ni/waste Marble Powder Catalyst Promoted by Transition Metals. Waste Management, 132, 96-104.

9. Li, B., Magoua Mbeugang, C. F., Huang, Y., Liu, D., Wang, Q., Zhang, S., 2022. A Review of Cao Based Catalysts for Tar Removal During Biomass Gasification. Energy, 244, 123172.

10. Shen, Y., Yoshikawa, K., 2013. Recent Progresses in Catalytic Tar Elimination During Biomass Gasification or Pyrolysis - A Review. Renewable and Sustainable Energy Reviews, 21, 371-392.

11. Bilgiç, G., Bendeş, E., Öztürk, B., Atasever, S., 2023. Recent Advances in Artificial Neural Network Research for Modeling Hydrogen Production Processes. International Journal of Hydrogen Energy, 48(50), 18947-18977.

12. Bilgiç, G., Öztürk, B., Atasever, S., Şahin, M., Kaplan, H., 2023. Prediction of Hydrogen Production by Magnetic Field Effect Water Electrolysis Using Artificial Neural Network Predictive Models. International Journal of Hydrogen Energy, 48(53), 20164-20175.

13. Qi, J., Zhang, K., Hu, M., Xu, P., Huhe, T., Ling, X., Yuan, H., Wang, Y., Chen, Y., 2023. Study on Waste Tire Pyrolysis Product Characteristics Based on Machine Learning. Journal of Environmental Chemical Engineering, 11(6), 111314.

14. Lei, C., Deng, J., Cao, K., Xiao, Y., Ma, L., Wang, W., Ma, T., Shu, C., 2019. A Comparison of Random Forest and Support Vector Machine Approaches to Predict Coal Spontaneous Combustion in Gob. Fuel, 239, 297-311.

15. Potnuri, R., Rao, C.S., Surya, D.V., Kumar, A., Basak, T., 2023. Utilizing Support Vector Regression Modeling to Predict Pyro Product Yields from Microwave-Assisted Catalytic Co-pyrolysis of Biomass and Waste Plastics. Energy Conversion and Management, 292, 117387.

16. Breiman, L., 2001. Random Forests. Machine Learning, 45(1), 5-32.

17. Yu, P.-S., Yang, T.-C., Chen, S.-Y., Kuo, C.-M., Tseng, H.-W., 2017. Comparison of Random Forests and Support Vector Machine for Real-Time Radar-Derived Rainfall Forecasting. Journal of Hydrology, 552, 92-104

18. Smola, A.J., Schölkopf, B., 2004. A Tutorial on Support Vector Regression. Statistics and Computing, 14(3), 199-222.
19. Awad, M., Khanna, R., 2015. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Springer Natures, 268.
20. Mutlu, A. Y., Yucel, O., 2018. An Artificial Intelligence Based Approach to Predicting Syngas Composition for Downdraft Biomass Gasification. Energy, 165, 895-901.
21. Elmaz, F., Yücel, Ö., Mutlu, A.Y., 2020. Predictive Modeling of Biomass Gasification with Machine Learning-based Regression Methods. Energy, 191, 116541.
22. Leng, E., He, B., Chen, J., Liao, G., Ma, Y., Zhang, F., Liu, S., E, J., 2021. Prediction of Three-phase Product Distribution and Bio-oil Heating Value of Biomass Fast Pyrolysis Based on Machine Learning. Energy, 236, 121401.
23. Xing, J., Luo, K., Wang, H., Fan, J., 2019. Estimating Biomass Major Chemical Constituents from Ultimate Analysis Using a Random Forest Model. Bioresource Technology, 288, 121541.
24. Irfan, M., Li, A., Zhang, L., Ji, G., Gao, Y., Khushk, S., 2021. Hydrogen-rich Syngas from Wet Municipal Solid Waste Gasification Using Ni/waste Marble Powder Catalyst Promoted by Transition Metals. Waste Management, 132, 96-104.
25. Choi, Y.-K., Cho, M.-H., Kim, J.-S., 2015a. Steam/oxygen Gasification of Dried Sewage Sludge in a Two-stage Gasifier: Effects of the Steam to Fuel Ratio and Ash of the Activated Carbon on the Production of Hydrogen and Tar Removal. Energy, 91, 160-167.
26. Kargbo, H.O., Zhang, J., Phan, A.N., 2023. Robust Modelling Development for Optimisation of Hydrogen Production from Biomass Gasification Process Using Bootstrap Aggregated Neural Network. International Journal of Hydrogen Energy, 48(29), 10812-10828.
27. Luo, S., Xiao, B., Hu, Z., Liu, S., Guo, X., He, M., 2009. Hydrogen-rich Gas from Catalytic Steam Gasification of Biomass in a Fixed Bed Reactor: Influence of Temperature and Steam on Gasification Performance. International Journal of Hydrogen Energy, 34(5), 2191-2194.
28. Nahil, M.A., Wang, X., Wu, C., Yang, H., Chen, H., Williams, P.T., 2013. Novel Bi-functional Ni-Mg-Al-Cao Catalyst for Catalytic Gasification of Biomass for Hydrogen Production with in Situ CO2 Adsorption. RSC Advances, 3(16), 5583.
29. Luo, S., Xiao, B., Hu, Z., Liu, S., Guo, X., He, M., 2009. Hydrogen-rich Gas from Catalytic Steam Gasification of Biomass in a Fixed Bed Reactor: Influence of Temperature and Steam on Gasification Performance. International Journal of Hydrogen Energy, 34(5), 2191-2194.
30. Barontini, F., Frigo, S., Gabbrielli, R., Sica, P., 2021. Co-gasification of Woody Biomass with Organic and Waste Matrices in a Down-draft Gasifier: An Experimental and Modeling Approach. Energy Conversion and Management, 245, 114566.
31. Li, B., Yang, H., Wei, L., Shao, J., Wang, X., Chen, H., 2017. Hydrogen Production from Agricultural Biomass Wastes Gasification in a Fluidized Bed with Calcium Oxide Enhancing. International Journal of Hydrogen Energy, 42(8), 4832-4839.
32. Gao, N., Liu, S., Han, Y., Xing, C., Li, A., 2015. Steam Reforming of Biomass Tar for Hydrogen Production over NIO/ceramic Foam Catalyst. International Journal of Hydrogen Energy, 40(25), 7983-7990.
33. Yusup, S., Khan, Z., Ahmad, M.M., Rashidi, N.A., 2014. Optimization of Hydrogen Production in In-situ Catalytic Adsorption (ICA) Steam Gasification Based on Response Surface Methodology. Biomass and Bioenergy, 60, 98-107.
34. Wei, L., Xu, S., Liu, J., Liu, C., Liu, S., 2008. Hydrogen Production in Steam Gasification of Biomass with CaO as a CO2 Absorbent. Energy & Fuels, 22(3), 1997-2004.
35. Faki, E., Üzden, Ş.T., Seçer, A., Hasanoğlu, A., 2022. Hydrogen Production from Low Temperature Supercritical Water CO-Gasification of Low Rank Lignites with Biomass. International Journal of Hydrogen Energy, 47(12), 7682-7692.

36. Ozbas, E.E., Aksu, D., Ongen, A., Aydin, M.A., Ozcan, H.K., 2019. Hydrogen Production Via Biomass Gasification, and Modeling by Supervised Machine Learning Algorithms. International Journal of Hydrogen Energy, 44(32), 17260-17268.

37. Balsora, H.K., Kartik, S., Joshi, J.B., Sharma, A., Chakinala, A.G., 2023. Artificial Neural Network-based Models for the Prediction of Biomass Pyrolysis Products from Preliminary Analysis. Industrial & Engineering Chemistry Research, 62(36), 14311-14319.

38. Chen, Y., Wang, Z., Lin, S., Qin, Y., Huang, X., 2023. A Review on Biomass Thermal-oxidative Decomposition Data and Machine Learning Prediction of Thermal Analysis. Cleaner Materials, 9, 100206.

39. Onsree, T., Tippayawong, N., 2021. Machine Learning Application to Predict Yields of Solid Products from Biomass Torrefaction. Renewable Energy, 167, 425-432.